Date: 6-3-99 Express Mail Label No. EL290991956US

Inventors: Philip P. Carvey and William J. Dally

Attorney's Docket No.: AVI99-02

PHOTONIC SWITCH USING TIME-SLOT INTERCHANGE

BACKGROUND OF THE INVENTION

High-speed telecommunications switches accept optical inputs and produce optical outputs but internally involve almost entirely electronic switching. Single-mode optical fibers are the technology of choice for long-haul transmission of information because they have very wide bandwidth, low attenuation, and low dispersion, making it possible to transmit information at very high bit rates (10^{10} b/s) over long distances (10^5 m) without the need of repeaters. Digital electronics, however, is the technology of choice for switching. Digital integrated circuits can switch connections in less than a nanosecond, and 10^5 to 10^6 digital gates can be incorporated on a single integrated circuit facilitating construction of the logic that controls a fast switch on a cycle by cycle basis.

Figure 1 shows a block diagram of a prior art telecommunications switch. Inputs arrive on single-mode optical fibers 101. In a typical application, these fibers carry a serial bit stream at 2.5Gbits/s (OC48) and are formatted using SONET (synchronous optical network) framing. The signal on each input fiber 101 is converted to electrical form by optical-to-electrical (O/E) converters 102. These converters demultiplex the data stream, converting the 2.5Gbits/sec serial stream into a 16-bit wide stream at 156Mbits/s. This electrical version of the input stream 103 is then input to an electronic switch 1-4. The switching fabric of the internet routers described in published PCT patent application number PCT/US98/16762 is an example of such a switch. The switch extracts individual packets or cells (depending on the protocol) from

the SONET frames (or other framing) of the incoming streams on each input and forwards each packet or cell to the output to which it is addressed. At the output, the packets or cells are encapsulated in a SONET frame and output electronically. This electrical output stream 105 is then output to an electrical-to-optical converter (E/O) and the optical output stream 107 drives the long-haul fiber to the next telecommunications switch.

It is attractive to use an electronic switch to direct optical signals because it is very easy to build logic and memory electronically, and very difficult to realize these functions optically. The switch requires logic to examine the contents of packets and determine where they are to be routed, and the switch requires memory to buffer packets according to a quality-of-service.

SUMMARY OF THE INVENTION

As the demand for bandwidth grows exponentially, both the number of ports on telecommunications switches, and the bandwidth per port also increases exponentially. These trends make it increasingly difficult to build electronic switches. Each electronic signal in a typical switch has limited bandwidth (usually less than 1 Gbits/s and often only100-200Mbits/s), thus wide paths are needed to carry the electronic form of an optical data stream. These wide paths add considerable expense to an electronic switch. As switches move to OC192 (10Gbits/s), 32-bit paths at 311Mbis/s per signal are needed to carry input electronically, and at OC768 (40Gbits/s), a 128-bit electronic bus at 311Mbits/s will be needed to carry the information on a signal optical fiber. Also, electronic signals can travel only a limited distance due to attenuation. This further complicates the design of electronic switches, particularly as the number of ports increases.

While an all-optical switch addresses the problem of electronic bandwidth, it has difficulty performing the logic needed to route packets or providing the memory

required for packet buffering. Also, optical switches typically switch very slowly (tens of microseconds to hundreds of milliseconds), making it impossible to switch individual packets or cells that may be less than 10 nanoseconds long (e.g., a 40-byte packet at OC768 lasts only 8ns).

In accordance with the present invention, a telecommunications switch comprises a plurality of optical inputs and a plurality of optical outputs. An optical switch operates with a schedule that is not directly determined by the input stream. The ordering units rearrange the order of data units, such as packets or cells, within data streams to correspond to the schedule of the switch. The reordering may be made in either input streams or output streams.

Preferred switches include a crossbar or a multi-stage interconnection network. The preferred reordering unit is a time-slot interchanger which contains a plurality of FIFOs. The FIFOs are implemented as circular buffers in a single dual port memory. The switch schedule may be fixed and balanced, or it may be unbalanced. In the latter case, the switch schedule may be determined by the average load between inputs and outputs. More specifically, the switch schedule may be determined by the number of data units queued from each input for each output in time-slot interchangers.

The present invention combines the best features of electronics and optics to overcome the bandwidth bottleneck of electronic switching, while at the same time using electronics to provide logic and memory. Also, this arrangement handles the switching of short (nanosecond) packets using optical switches that can be reconfigured at microsecond to millisecond time scales.

The present invention does require conversion of the input signal from electrical to optical and back again. However, this conversion is required in any case for signal regeneration.

All of the operations performed electronically on the data streams are local to a single input, and thus require a minimum of wide electrical pathways. Electronics are used to provide logic and memory, and optics are used to provide switching bandwidth. Electronic memory is used to buffer packets or cells until they are to be transmitted. All of the switching is performed optically.

The present invention is also well matched to the properties of an optical switch. By grouping many packets traveling to the same output together in time, the present invention allows short (nanosecond) length packets to be switched while reconfiguring the optical switch only once every sub-frame, frame, or multi-frame time period (microseconds or milliseconds). Also, the optical switch can be controlled with a fixed pattern, without the need to examine the arriving data to configure the switch. Alternatively, the switch can be controlled with an adaptive pattern that uses only average input to output load statistics to balance load across the switch. In either case, the configuration of the optical switch does not directly depend upon the data it is switching.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings in which like reference characters refer to the same parts throughout the different views. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

Fig. 1 is an illustration of a conventional telecommunications switch;

Fig. 2 is an illustration of one embodiment of the present invention;

Fig. 3 is an illustration of an alternative embodiment of the invention;

Fig. 4 shows an example of the schedule of input/output connections in a switch;

Fig. 5 illustrates the connections of Fig. 4 in graphical form;

Fig. 6 illustrates an exemplary input sequence of data packets in the reordering of the sequence;

Fig. 7 illustrates an unbalanced schedule;

Fig. 8 illustrates a FIFO implementation of a time-slot interchanger;

Fig. 9 illustrates a dual port memory implementation of a time-slot interchanger.

DETAILED DESCRIPTION OF THE INVENTION

The present invention operates by converting the incoming streams to electronic form, but performs no switching on these electronic streams. As illustrated in Fig. 2, each electronic input stream 103 is input to a time-slot interchanger (TSI) 111. The TSI examines the packets within an input stream, using logic to make routing decisions, and reorders them in time (exchanging their time slots) so that packets traveling to a given output are grouped together in time. The reordered streams 112 are then converted back to optical form. These reordered optical streams 113 are then input to an all-optical switch that changes its connectivity in a fixed pattern at a relatively slow rate. Each TSI schedules the packets entering the optical switch to arrive at a time when its input is connected to the packet's destination.

Fig. 6 shows an input stream of packets 201 on an input line 101 and a reordered stream of packets 202 on internal line 113. Each packet is marked with a number denoting the output port to which it must be forwarded and a letter denoting its sequence. For example, packet 2b is the second packet that must be forwarded to output port 2.

The TSI 111 reorders the packets in input stream 201 so that all packets destined for the same output are consecutive and occupy a fixed time period on the reordered stream 202 on internal line 113. During time period t=1, packets destined to output port 1 are transmitted on line 113, during period t=2, packets destined to port 2 are on line

113, and so on. This ordering of packets matches the connections provided by the optical switch 114 under control of sequencer 115. During time period t=1, the switch connects input 1 to output 1, during time period t=2, the switch connects input 1 to output 2, and so on. Thus, by reordering the packets in time, the TSI 111 is able to route the packets by lining them up with periods during which the switch will be connected to the desired output. For example, the packet labeled 2b is placed on input 1 of the switch 114 during period t=2 and thus is forwarded to output 2 of the switch as desired.

The details of the switch and sequencer 115 operation are illustrated in Figs. 4 and 5. Fig. 4 shows, in tabular form, how the sequencer directs the switch to connect inputs to outputs during each of the four periods. Fig. 5 shows this same information graphically. During period t=1, for instance, the figures show that the inputs are connected straight across to the outputs. During period t=2, the connections are rotated by 1 with input 1 connected to output 2, input 2 to output 3, and so on. In general, during period t=x, input i is connected to output i+x-1 (mod 4). The sequencer repeats these four connections indefinitely, connecting each input to each output in turn. The interchangers, knowing this pattern in advance, schedule packets to appear on a switch input during the time period when that input is connected to the desired output. In a preferred embodiment of the present invention, each time period or time-slot is 10 microseconds in length.

The TSI 111 that reorders packets to make packets destined for the same output contiguous can be conceptually implemented using a FIFO for each output as illustrated in Fig. 8 for the case of four outputs. The interchanger consists of four FIFOs 151-154, and a multiplexer, 160. As each packet arrives on electronic input stream 103, it is examined to determine the output port to which it should be forwarded. It is then appended to the end of a FIFO queue of packets associated with this output. During each time-slot of the optical switch, multiplexer 160 is switched to select the FIFO

associated with the selected output of the switch. Packets are then read from this selected FIFO to reordered electronic input stream 112 until the time-slot is over or the FIFO is emptied. If the FIFO is emptied before the time-slot is over, idle symbols are transmitted through the switch. If the time-slot expires before the FIFO is empty, the packets remaining in the FIFO are retained and will be transmitted during a later time-slot when the switch is connected to the same output.

As an example of interchanger operation, consider the input stream 201 shown in Fig. 6. When packet 1a arrives, it is appended to FIFO 151 (associated with output 1), next packet 2a is appended to FIFO 152, packet 1b is appended to 151 behind 1a, and so on. During time-slot t=1, when this switch input is connected to output 1, the multiplexer 160 selects FIFO 151 and packets 1a, 1b and 1c are output in sequence to reordered input stream 112 to be forwarded through the switch to output 1. (This assumes that no earlier packets remained in the FIFO to be transmitted before these packets.) Next, during time-slot t=2, FIFO 152 is selected and packets 2a and 2b are transmitted on stream 112 to be forwarded to output 2, and so on.

In a preferred embodiment of the present invention, the four FIFOs 151-154 of the TSI 111 are implemented with a single dual port memory 221, as illustrated in Fig. 9. In this embodiment, the contents of each FIFO are stored in dedicated circular buffers in memory 221 (denoted by the dotted lines). Memory 221 is a dual port memory with one read port and one write port. The write port consists of data input 222 and write address 223. The read port consists of data output 224 and read address 225. Each circular buffer is indexed by one of the tail pointers 231-234 and one of the head pointers 251-254. A write address multiplexer 241 selects one of the tail pointers to be used as the write address for memory 221. Similarly, a read address multiplexer 261 selects one of the head pointers to be used as the read address for memory 221.

When a packet arrives at the TSI of Fig. 9, the packet is examined to determine its output port and multiplexer 241 selects the tail pointer associated with this output. As each word of the packet arrives, it is placed on the data input line 222 and written into the appropriate circular buffer using the selected tail pointer as the write address 223. The tail pointer is then incremented with a limit check to wrap addresses within the circular buffer. The tail pointer is also compared to the corresponding head pointer to check for a buffer-full condition. When a new time-slot begins, the output multiplexer selects the head pointer associated with the output corresponding to this time-slot to be used as the read address. All of the packets associated with the output are then read out of the selected circular buffer, incrementing the head pointer (with circular buffer wrapping) after each word of a packet is read. The head pointer is compared to the corresponding tail pointer after each increment to detect an empty circular buffer.

In one embodiment of the present invention, optical switch 114 is implemented using a $LiNbO_3$ non-linear optical crossbar switch. In an alternate embodiment, the switch is implemented as a multi-stage optical switching network, as described in Chamberlain, et al., "Design of an Optically-Interconnected Multiprocessor", IEEE 0-8186-8572-7/98.

The present invention allows data to be switched to the desired output of an optical switch with fixed control (the data need not be examined to control the switch). However, it can only achieve full throughput when each input carries an equal amount of traffic destined for each output. This is because the switch schedule shown in Figs. 4 and 5 is balanced, with each input connected to each output for equal amounts of time. If the input traffic is not balanced, one or more of the FIFOs in the interchanger may be overrun, resulting in loss of data.

This requirement for balanced traffic can be overcome by using an unbalanced schedule, as illustrated in tabular form in Fig. 7. This figure illustrates an eight period schedule. In this schedule, input 1 sends three units of traffic to output 2, one unit of traffic to output 3, and two units of traffic to all other outputs. Input four sends three units to output 3, one unit to output 2, and two units to outputs 1 and 4. The traffic from inputs 2 and 3 is balanced.

Unbalanced switch schedules such as the one shown in Fig. 7 can be generated automatically by examining the occupancy of each of the four FIFOs in each of the four interchangers. When the occupancy of a FIFO (in the example, the FIFO from input 1 containing traffic destined for output 2) exceeds a threshold, the interchanger finds the input 1 FIFO with the least occupancy (in this case, the one containing traffic to output 3), and swaps outputs with the interchanger (in this case, the input 4 interchanger) that connects to 2 during the cycle that 1 would normally connect to 3. Hence, the schedule shown in Fig. 7 is generated.

While this schedule is dependent on the data being transported, it still has two properties that make it suitable for driving an optical switch. First, it reconfigures the switch at a frequency much lower than the packet rate. The switch is still set to one configuration during a period that spans many packets. In this case, however, the periods may be unbalanced to match the unbalance in input traffic. Second, it is latency sensitive. If the interchanger FIFOs are sufficiently long, a considerable period may pass between when an imbalance is detected and when the switch schedule is changed.

In an alternate embodiment of the present invention, illustrated in Fig. 3, the position of the optical switch 114 and the TSI 111 are reversed. In this case, the switch output taken by a packet on the present telecommunications switch is determined by the TSI on the output of the upstream telecommunications switch: the switch that drives the

input line 101. The TSI 111 on the output of that switch schedules packets so that they appear during the proper time-slot of the present switch.

One skilled in the art will understand that the present invention can be realized in a number of different forms. For example, the optical switch may be implemented using thermally actuated directional couplers, mechanically actuated optical switches, or piezoelectrically driven optical switches. The configuration of the switch may also be varied. While we have described the present invention in the context of a four-port crossbar switch, a crossbar with any number of ports may be implemented, a multi-stage network may be implemented, or a direct interconnection network may be implemented. The present invention can also be implemented with time slots of varying sizes. With slow acting mechanical, thermal, or piezoelectrically-driven switches, the time-slot may be increased to a millisecond or more (at the expense of larger FIFO buffers in the TSI).

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended claims.